Estimating the Difference between Published and Unpublished Effect Sizes: A Meta-Review

Joshua R. Polanin, Ph.D.

Emily E. Tanner-Smith, Ph.D.

Emily A. Hennessy, MPhil.

Vanderbilt University

Author Notes

Joshua R. Polanin, Emily E. Tanner-Smith, & Emily A. Hennessy, Peabody Research Institute, Vanderbilt University. Research for the current study was partially supported by Loyola University Chicago, Advanced Dissertation Fellowship as well as an Institute of Education Sciences Postdoctoral Training Fellowship grant to Peabody Research Institute, Vanderbilt University (R305B100016). Opinions expressed herein do not necessarily reflect those of Loyola University Chicago or the Institute of Education Sciences. Correspondence regarding this article should be addressed to Joshua R. Polanin, Peabody Research Institute, 230 Appleton Place, Nashville, TN 37203.

Email: jpolanin@dsgonline.com

Citation.
Polanin, J.R., Tanner-Smith, E.E., & Hennessy, E.A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*, 207 – 236. doi:10.3102/0034654315582067

Abstract

Practitioners and policymakers rely on meta-analyses to inform decision-making around the allocation of resources to individuals and organizations. It is therefore paramount to consider the validity of these results. A well-documented threat to the validity of research synthesis results is the presence of publication bias, a phenomenon where studies with large and/or statistically significant effects, relative to studies with small or null effects, are more likely to be published (Rothstein, Borenstein, & Sutton, 2005). We investigated this phenomenon empirically by reviewing meta-analyses published in top tier journals between 1986-2013 that quantified the difference between effect sizes from published and unpublished research. We reviewed 383 meta-analyses of which 81 had sufficient information to calculate an effect size. Results indicated that published studies yielded larger effect sizes than those from unpublished studies ($\bar{d}$ = 0.18, 95% CI [0.10, 0.25]). Moderator analyses revealed that the difference was larger in meta-analyses that included a wide range of unpublished literature. We conclude that intervention researchers require continued support to publish null findings and that meta-analyses should include unpublished studies to mitigate the potential bias from publication status.

Keywords: publication bias, meta-review, effect size, meta-analysis.

Estimating the Difference between Published and Unpublished Effect Sizes: A Meta-Review

Practitioners and policymakers rely on high quality primary research as well as meta-analyses to inform decision-making around the allocation of resources and provision of services to individuals and organizations. This paradigm extends to the results of studies both null and significant because ensuring the discontinuation of ineffective, or worse, harmful programs, is as important as ensuring effective programs are distributed widely. Greenwald (1975), in a landmark study on the "prejudice against the null hypothesis" (p. 1), posited that null research findings are often interpreted as a failing of the researcher (or design), instead of the probable hypothesis that successfully detecting a significant effect is much less likely. Numerous studies have shown, however, that across the peer-reviewed, published literature detecting a statistically significant effect is often the rule rather than exception (Ionnidis, 2005; Ioannidis & Trikalinos, 2007). As a result, there have been calls recently by journals and researchers alike to include null findings from high quality studies (Cumming, 2013).

Despite these well-intentioned calls to action, statistically significant research remains the norm in peer-reviewed, published research (Rothstein, Borenstein, & Sutton, 2005). Indeed, prior research has documented that (a) studies with non-significant findings are often left unpublished (Ferguson & Heene, 2012), (b) authors have a tendency to report results for outcomes with statistically significant effects (Pigott, Valentine, Polanin, Williams, & Canada, 2013), and (c) external referees tend to provide more favorable peer reviews to studies with large or statistically significant effects (Lee, Sugimoto, Zhang, & Cronin, 2013). In addition, favorable outcomes receive quicker time to publication (Ioannidis, 1998) and a higher rate of citation (Tanner-Smith & Polanin, 2014). The totality of these biases is often referred to as dissemination bias.

The proliferation of dissemination biases has the potential to impact not only the validity of meta-analytic results, but those of primary research as well. Dissemination biases create the illusion of theory confirmation, potentially leading to the continuation of programs or policies that are ineffective or, worse, harmful. Moreover, continuation of funds to ineffective programs inhibits the growth of potentially new and important research. At a minimum, dissemination bias in a particular field may cause an overestimation of potential future effects, which has an impact on researchers' preliminary power and cost analyses (Rothstein et al., 2005). The recent movement to represent clinical trials or funded research in data repositories, such as *ClinicalTrials.gov*, addresses this concern by mandating that primary researchers publish a study's results, regardless of the study's findings.

Erroneous conclusions or misleading results reporting are particularly problematic for meta-analyses that synthesize evidence on the effectiveness of educational or social interventions. Fortunately, most authors of research syntheses are aware of the dangers of publication bias (Ferguson & Brannick, 2012), and there are several analytic methods available for meta-analysis authors to assess the possibility and potential impact of publication bias on their review findings (Sutton, 2009). One common method is to test for differences in mean effect sizes as a function of publication status—that is, asking whether the observed effects are larger or smaller in published versus unpublished studies included in the meta-analysis (Lipsey, 2009). This empirical test addresses whether excluding unpublished literature would have changed the meta-analytic findings, and is generally used to elucidate the potential dangers of having only including published literature in the meta-analysis.

The purpose of this meta-review is to capitalize on these tests of average effect size differences by publication status. Our primary research question is: What is the observed difference in mean effect sizes for published versus unpublished studies, as reported in meta-analyses in the fields of education and psychology? The project's results provide an estimate of possible dissemination bias, and, moreover, have the potential to inform both primary researchers and meta-analysts of the phenomenon of dissemination bias within the social science literature. We chose to focus this meta-review on education and psychology because less attention has been paid to publication bias in the social sciences relative to health and medical sciences (Rothstein, et al., 2005).

Research on Publication Bias

Systematic literature reviews aim to locate and include all relevant studies on a particular topic using transparent procedures designed to maximize replicability and minimize potential biases. The inclusion of all relevant studies that meet pre-specified eligibility criteria means that studies should be included in a review regardless of publication status, and thus most systematic reviews attempt to include studies controlled by academic/commercial publishers as well as grey literature. Grey literature refers broadly to any study "which is produced on all levels of government, academic, business, and industry in print and electronic formats, but which is not controlled by commercial publishers" (Hopewell, Clarke, & Mallett, 2005, p. 49). Henceforth, we refer to any study report outside of commercially published journal articles as unpublished grey literature. Dissertations and theses are one of the most common type of grey literature included in systematic reviews, but other forms of grey literature may include conference abstracts/presentations, books and book chapters, unpublished technical reports or white papers, and unpublished datasets received from primary study authors.

Dissemination and publication biases. The primary reason meta-analysts deliberate and advocate for the inclusion of grey literature in systematic reviews is publication bias, or the higher likelihood for studies with large or statistically significant findings to be published relative to studies with small or null effects (Rothstein et al., 2005). Publication bias is one form of bias under the larger umbrella of dissemination bias, which refers to the phenomenon whereby the selective reporting, selective publication, and/or selective inclusion of scientific evidence in a systematic review yields a biased answer, on average (Bax & Moons, 2011). Other dissemination biases, for instance, might include time-lag bias (studies with large/significant results are published more quickly), citation bias (studies with large/significant results are cited more frequently), and outcome-reporting bias (outcomes with large/significant results are more likely to be reported by primary study authors).

Numerous empirical studies have documented the existence of dissemination biases, mostly in the medical sciences. For instance, Easterbrook, Berlin, Gopalan, and Matthews (1991) surveyed projects approved by the Central Oxford Research ethics committee in the 1980s, and found that studies with statistically significant findings were more likely to be published. Similarly, Decullier, Lheritier, and Chapuis (2005) retrospectively investigated studies approved by French research ethics committees in the 1990s and found that studies with significant and positive results were more likely to be published relative to those with inconclusive or non-confirmatory findings. Chan, Hróbjartsson, Haahr, Gøtzsche, and Altman (2004) reviewed protocols of randomized controlled trials conducted in Denmark and found that whereas 71% of statistically significant outcomes were reported, only 56% of non-significant outcomes were reported.

Reviews of medical meta-analyses have yielded similar findings. Hopewell, Loudon, Clarke, Oxman, and Dickersin (2009) reviewed Cochrane reviews to investigate the association between publication status and research findings for registered clinical trials and found that trials with statistically significant results were more likely to be published. Dwan and colleagues (2008) examined systematic reviews of medical interventions and found that 40-62% of studies "had at least one primary outcome that was changed, introduced, or omitted" (p. 1). A recent review of 51 meta-analyses across the communication sciences literature (Levine, Kelli, & Carpenter, 2009) showed that effect sizes were negatively correlated with sample sizes. Their findings indicated that statistically significant findings were more likely to be published especially among smaller sample studies (instead of small sample studies published regardless of effect), and the authors concluded that resulting meta-analyses would likely overestimate the true population effect. Thus, the medical and communication sciences literature has a well-documented trend toward publishing statistically significant and large treatment effects (Rothstein et al., 2005).

Despite consistent evidence of publication bias in the medical literature, Torgersen (2006) asserted that few studies have examined its presence in the educational or psychological sciences. Of the studies conducted, for example, Smith's (1980) review of 12 meta-analyses from education found that, on average, studies published in peer-reviewed journals had a mean effect size one-third of a standard deviation larger than unpublished studies. Kulik and Kulik (1989) found similar results when examining their own four meta-analyses in computer-based instruction in elementary and secondary levels, ability grouping, and mastery learning systems. The authors compared the effect sizes between published and unpublished literature (dissertations) and found evidence of publication bias in two of the four meta-analyses. They concluded that publication bias was only one possible factor, however, because dissertation authors were less experienced researchers and this could explain the lack of publication. Lipsey and Wilson (1993) also examined publication bias in education by reviewing 302 meta-analyses of education-based studies. Of the 302 meta-analyses reviewed, only six reported effect sizes that were negative and more than 85% of studies reported effect sizes larger than 0.2 (in the standardized mean difference metric). More recently, Pigott et al. (2013) investigated outcome-reporting bias in education by tracking dissertations after they were later published in journals. Of the 79 dissertation studies that were subsequently matched with a later journal publication, 36 appeared to experience some form of outcome censoring due to statistical significance. The non-statistically significant outcomes were 30% less likely to appear in publication compared to statistically significant outcomes. Clearly, publication bias can have an impact on the validity of meta-analyses and their conclusions.

Methods for assessing potential publication bias. Several analytic methods have been proposed to assess the potential presence and impact of publication bias in a meta-analysis (Ferguson & Brannick, 2012; Ioannidis, Munafo, Fusar-Poli, Nosek, & David, 2014). These techniques vary in complexity (Hedges & Vevea, 2005) and rigor (Rosenthal, 1991), although some techniques tend to be more popular with review authors (Ahn, Ames, & Myers, 2012). One visual analysis technique for assessing the possibility of publication bias in a meta-analysis is an inspection of a funnel plot (Sterne, Becker, & Egger, 2005). Funnel plots are scatter plots of effect size estimates (typically shown on the x-axis) and some measure of precision of the effect size estimate (typically the inverse of the standard error or sample size, shown on the y-axis). In the absence of publication bias, one expects the effect sizes to be dispersed evenly throughout the graph and to be symmetric around the mean effect size. If studies with smaller sample sizes and small/null/negative effect sizes are not included in the meta-analysis, the funnel plot appears either asymmetric or the bottom half plot empty. There are numerous problems with this visual method for assessing potential publication bias; foremost of which is that it relies on a subjective visual assessment of the a/symmetry of the plot, or that plot asymmetry could be due to numerous factors other than publication bias (Schild & Voracek, 2014).

Due to the limitations of visually inspecting funnel plots, other methods have been developed to attempt to quantify the asymmetry in the funnel plot. For example, regression-based approaches attempt to estimate the magnitude of association between effect size estimates and their precision and include such tests as Egger, Peters, and Harbord tests for funnel plot asymmetry (e.g., Egger, Smith, Schneider, & Minder, 1997; Harbord, Egger, & Sterne, 2006; Peters, Sutton, Jones, Abrams, & Rushton, 2006; Rücker, Schwarzer, & Carpenter, 2008). Alternatively, the trim and fill method (Duval & Tweedie, 2000) is another commonly used method that estimates the number of missing studies in a meta-analysis, assuming a symmetric funnel plot should exist. The procedure does so by assessing funnel plot asymmetry and then deducing the number of potentially missing studies by attempting to produce a symmetric funnel plot. As a result, one may calculate an average effect size that includes imputed missing effect sizes and compare this value to the original effect size generated without imputed values.

Rosenthal (1991) and Orwin (1983) proposed various measures of "file-drawer" or "failsafe N" statistics, which estimate the number of presumably missing studies that could be added to a meta-analysis before it yielded a null mean effect size. This statistic has recently fallen out of favor with methodologists because it can be easily misrepresented and null findings may not always produce effect sizes close to zero (Becker, 2005; Higgins & Green, 2011). The p-curve is another method that attempts to uncover selective reporting, or "p-hacking," in primary reports (Simonsohn, Nelson, Leif, & Simmons, 2013). Although all of the aforementioned techniques are useful for assessing the possibility of publication bias in a meta-analysis, they typically do not focus on the observed difference in mean effect sizes for published and unpublished studies' effect sizes.

As noted previously, one of the most common methods for assessing the possibility of publication bias in a meta-analysis is to test for the differences in mean effect sizes across the published versus unpublished studies included in the review (Polanin & Pigott, 2014). This test is commonly conducted as a simple moderator analysis to examine whether the observed difference in mean effect sizes across the published and unpublished studies is larger than would be expected due to chance. When comparing means across two groups (i.e., published vs. unpublished studies), this moderator test can be converted to an effect size. Consideration of more than two groups, for example, when the purpose is to test for differences between multiple types of grey literature (e.g., journal articles, books, dissertations), results in the authors utilizing a *Q*-between test (analogous to an F-test in an ANOVA model). Either the result of the *Q*-between or the z-score test provides a measure of the magnitude of the observed difference in mean effect sizes across the publication status groups.

The Current Study

This meta-review synthesizes results across meta-analyses in the fields of education and psychology, in an attempt to capture the average magnitude of difference in mean effect sizes across published and unpublished studies. The purpose of this meta-review is to capitalize on the moderator tests of differences in mean effect sizes by publication status. Our primary research question is as follows: What is the observed difference in mean effect sizes for published versus unpublished studies, as reported in meta-analyses in the fields of education and psychology?

Our secondary research question is whether any characteristics of the meta-analyses themselves may influence the magnitude of observed differences in mean effect sizes between published and unpublished studies.

Method

Search and Screen

The purpose of this meta-review is to estimate the average difference in the magnitude of effect sizes between published and unpublished primary studies, as reported in existing meta-analyses in the fields of education and psychology. As such, we screened all published meta-analyses in the journals *Review of Educational Research* and *Psychological Bulletin.* These two journals were chosen to represent education and psychology because they publish high-quality meta-analyses and because they each consistently generate large impact factors. To be eligible for inclusion in this meta-review, the published research synthesis must have been conducted between 1986 – 2013 and must have presented quantitative synthesis results (i.e., conducted a meta-analysis). The nature of the meta-review also meant that the meta-analyses must have included unpublished, grey literature. There were no restrictions placed on the type of studies synthesized (e.g., meta-analyses could synthesize experimental intervention effectiveness research or non-experimental correlational research) or the method of synthesis. We were interested in producing a generalizable estimate of the difference in effect sizes and therefore purposively enabled the inclusion criteria to be broad.

Each citation's title and abstract was hand screened for inclusion. If a study appeared to meet these broad screening criteria, we retrieved the full-text version and screened the full article for final eligibility determination. Across the two journals, we screened a total of 1,858 citations; 383 citations appeared to meet the inclusion criteria and were retrieved for full-text screening. One researcher screened all 383 articles and another researcher randomly selected and screened 10% of the studies. The two reviewers agreed on the decision to include the meta-analysis often (96%) and the disagreements were resolved via consensus. Of the screened 383 articles, 81 met the inclusion criteria and had sufficient information available to calculate an effect size.

Coding

A comprehensive coding manual was created for this project and four major sections guided the coding process. The first section coded basic information about the meta-analysis: Author names, date of publication, the title of the article, the publication source, and whether the meta-analysis was funded. The second section detailed the meta-analyses' characteristics: A description of the synthesis, outcome construct type (educational achievement or psychological), effect size metric (standardized mean difference or correlation coefficient), and type of synthesis model (fixed-effect, random-effects, or both). The third major section detailed the types of grey literature included in each meta-analysis. On a few occasions, the meta-analysis authors failed to delineate the types of grey literature provided or simply stated that "unpublished studies" were included in the review. In these cases, we labelled this type of grey literature as "general." We also coded the total number of published and unpublished studies included in each meta-analysis. The fourth section of the coding manual detailed information needed to extract effect sizes from each meta-analysis.

Effect Size Extraction

The effect size of interest was the difference between the average effect sizes for published versus unpublished literature included in each meta-analysis. This can be represented as

$$Diff_i = T_{Published} - T_{Unpublished} \qquad \text{EQ1}$$

where $T_{Published}$ represented the average effect size from published literature and $T_{Unpublished}$ represented the average effect size from unpublished literature in any given meta-analysis. This effect size was calculated such that positive values indicated published studies had larger effects than unpublished studies. The standard error of this effect size is the sum of the standard errors for each average effect size. If the meta-analyses presented effect sizes for studies across multiple types of unpublished literature categories (e.g., dissertations, conference abstracts, etc.), we first combined the multiple sources into one category (i.e., all unpublished literature) before estimating the above effect size statistic.

Because meta-analyses reported the results of this moderator test in a variety of forms, we could not always directly calculate this simple effect size statistic. For example, a portion of the meta-analyses provided only the *Q*-between test statistic with 1 *df*. In this case, we converted the *Q*-between statistic with 1 *df* to the z-metric by taking its square root. As such, and in order to synthesize effects across all meta-analyses included in the sample, we transformed all effect sizes into the z-metric following Borenstein, Hedges, Higgins, and Rothstein (2009). For each meta-analysis *i* that provided information to calculate a difference estimate, we calculated a z-statistic:

$$z_i = \frac{Diff_i}{SE_i} \qquad \text{EQ2}$$

where $Diff_i$ is the difference calculated in equation (1) and the SE_i is the standard error for meta-analysis *i*. Again, this z-statistic effect size was coded such that positive values indicated published studies had larger mean effects than unpublished studies.

Finally, to guard against potential outcome reporting biases by review authors, we estimated the difference between published and unpublished studies directly from included summary tables when the authors reported that they conducted the moderator test but failed to provide empirical results. For instance, Elfenbein and Ambady (2002) indicated that a moderator test was conducted, but that the test yielded a non-significant result. The authors, in addition, provided a table of every study included in the meta-analysis as well as the effect size extracted. As a result, we extracted this summary table and directly calculated the effect size difference of interest.

Analysis

We conducted a random-effects meta-analysis, using robust variance estimation, to synthesize the weighted *z*-statistics estimated from each meta-analysis (Hedges, Tipton, & Johnson, 2010; Tanner-Smith & Tipton, 2014). The use of robust standard errors allows for the inclusion of multiple effect sizes per study. Similar to traditional meta-analysis, the weight of each study remained the inverse of the z-statistics' variance (i.e., the square of the standard error; Lipsey & Wilson, 2001). The result is an inverse-weighted, average *z* statistic. The standard error of the weighted average *z*, used to calculate confidence intervals, is the square root of the inverse of the sum of the weights. Congruent with standard meta-analytic procedures, we also estimated the heterogeneity among the meta-analyses via the τ^2 and I^2 statistics (Higgins & Thompson, 2002). To answer our first research question and provide an estimate of the magnitude of the difference between published and unpublished studies, we transformed the weighted average *z* back into the standardized mean-difference metric (Borenstein et al., 2009).

To answer our second research question, we also conducted moderator tests to determine if the effect sizes varied as a function of different subgroups of meta-analyses. Given the exploratory nature of these analyses, we chose to conduct these tests using a meta-regression framework, first by conducting univariate meta-regression models and then conducting a multivariable meta-regression model. We explored the following potential effect size moderators: effect size metric (standardized mean difference or correlation coefficient), outcome construct type (achievement or psychological outcome), whether the meta-analysis included dissertations as their only grey literature source, whether the estimated effect size derived from a *Q*-between model, whether the review received funding, the percentage of unpublished literature included in the meta-analysis, and the date of publication. Model fit was assessed using an adjusted R^2 (Cheung, 2013). All calculations and figures were prepared using the R package robumeta (Fisher & Tipton, 2014).

Results

Descriptive Results

This meta-review synthesized effect sizes from 81 meta-analyses, 27 meta-analyses derived from the *Review of Educational Research* and 54 from *Psychological Bulletin* (see Table 1). A majority of the meta-analyses used the standardized mean-difference effect size metric (k = 57, 70.4%) and the rest used the correlation coefficient (k = 24, 29.6%). Most effect sizes indexed effects for psychological outcomes (k = 47, 58.0%); others measured achievement outcomes (k = 30, 37.0%) and a few reported both types of outcomes (k = 4, 4.9%). About one-half of the meta-analyses reported that they received some type of funding (k = 44, 54.3%). A large percentage of meta-analyses failed to report the type of synthesis model utilized (k = 25, 30.9%) or used a random-effects analysis (k = 35, 43.2%); a few reported using a fixed-effect model (k = 10, 12.3%) or both fixed- and random-effects models (k = 11, 13.6%). Seven meta-analyses (8.6%) indicated that a moderator test was conducted and failed to provide relevant summary statistics, but included a summary table of the studies and effect sizes which could be used to estimate a difference statistic. Most of the meta-analyses were published after 2000 (k =

58, 71.6%). Figure 1 provides a descriptive graph of the number of meta-analyses published each year and includes how many of these meta-analyses reported conducting publication bias moderator analyses.

It is also informative to describe the quantity and types of grey literature included (Table 2). We should note, however, that to be included in this review, the meta-analyses must have included grey literature and conducted a test of the difference in effect size magnitude by publication status. Therefore, the percentages of included grey literature may not reflect the population of systematic reviews because not all systematic reviews include grey literature (Polanin & Pigott, 2014). For this sample of meta-analyses, 68.5% of the included primary studies were journal articles (Mean number of studies = 55.74, *SD* = 56.74) and 31.5% were some form of grey literature (*M* = 22.19, *SD* = 23.11). The most common type of grey literature was dissertation abstracts (*M* = 13.43, *SD* = 18.42). The next most commonly used grey literature type was general unpublished literature (*M* = 5.03, *SD* = 2.96); this category included unpublished datasets, government reports, documents retrieved from database searches but not published elsewhere, or a mix of unpublished articles where the authors did not distinguish the study type. Review authors included conference abstracts (*M* = 2.34, *SD* = 8.20), books (*M* = 0.56, *SD* = 2.94), and book chapters (*M* = 0.82, *SD* = 3.40) infrequently. Across all meta-analyses, 31.7% included dissertations/theses (hereafter referred to as dissertations) as their only source of grey literature.

Analytical Results

The 81 meta-analyses provided information to calculate 147 *z*-statistics that indexed the difference in mean effect sizes for published versus unpublished studies (Figure S1). The average study contributed more than one effect size (*M* = 1.80, SD = 2.05, Min = 1, Max = 14).

Synthesizing the *z*-statistics across all 147 effect sizes, using robust standard errors to account for within-study clustering, yielded a statistically significant average effect ($\bar{z}$ = 0.80, 95% CI [0.45, 1.15]), indicating that published studies had significantly larger mean effects than unpublished studies. We transformed the inverse-weighted, average z-statistic into the d-metric ($\bar{d}$ = 0.18, 95% CI [0.10, 0.25]). This indicates that on average, published studies reported effects that were 0.18 standard deviations larger than the effects reported in unpublished studies.

We observed a large amount of heterogeneity among the effect sizes (τ^2 = 5.86, I^2 = 99.8%). Therefore, we also conducted exploratory meta-regression analyses to examine whether the difference in mean effects between published and unpublished studies varied according to characteristics of the meta-analyses (Table 3). Only one of the six meta-regression models yielded traditional statistically significant differences (i.e., $p < .05$). Namely, meta-analyses that included dissertations as their only source of grey literature yielded a statistically significantly (b = -0.77, 95% CI [-1.51, -0.03]) lower average z-statistic ($\bar{z}$ = 0.28, 95% CI [-0.33, 0.90]) relative to studies that included all types of grey literature ($\bar{z}$ = 1.03, 95% CI [0.60, 1.45]). In other words, meta-analyses that only included dissertations as their grey literature source concluded that the difference between published and unpublished studies' (i.e., dissertations) effect sizes were not substantial. To account for covariance among the variables and further ensure the validity of these results, we also conducted a multiple predictor meta-regression model that included all of the hypothesized variables. The results again indicated that the only significant result derived from dissertations being used as the only source of grey literature (b = -0.83, 95% CI [-1.63, -0.03]).

Although no other meta-regression analyses yielded significant results, we believe it is important to highlight a few other results as potential areas for further research. For instance, results indicated that (b = 0.27, 95% CI [-0.42, 0.95]) meta-analyses receiving funding had a larger difference ($\bar{z}$ = 0.92, 95% CI [0.35, 1.49]) relative to unfunded studies ($\bar{z}$ = 0.65, 95% CI [0.26, 1.05]). One possible explanation for this is the amount and types of unpublished literature included in funded meta-analyses. Funded meta-analyses included more primary studies (Funded: M = 77.00, SD = 70.65; Unfunded: M = 70.50, SD = 52.85), and 32 of the 44 funded studies (72.7%) included diverse sources of grey literature (i.e., sources other than dissertations) whereas only 25 of the 38 unfunded meta-analyses (65.7%) included grey literature sources other than dissertations. Although this analysis was purely exploratory, the trend indicates that funded meta-analyses may include more diverse types of grey literature and thus yield a more conservative, yet potentially less biased, estimate of the overall effect size.

Discussion

The purpose of this meta-review was to summarize the observed differences in effect sizes between published and unpublished literature, as reported in meta-analyses in the educational and psychological sciences. After a comprehensive search, screen, and selection process of the citations from two leading review journals, we identified 81 meta-analyses and 147 effect sizes that were eligible and included in this meta-review. We coded and extracted effect sizes that represented the magnitude of difference in the mean effect sizes for published and unpublished studies included in the meta-analyses. Results from the meta-review indicated that, on average, published studies had mean effect sizes that were 0.18 standard deviations larger than the mean effect sizes for unpublished studies included in the meta-analyses (95% CI = 0.10, 0.25). Exploratory moderator analyses indicated that meta-analyses that included dissertations as their only grey literature source were significantly less likely to observe effect size differences between published and unpublished studies. Overall, these results provided evidence of the existence of publication bias in education and psychology.

Limitations

Several limitations should be considered. For instance, this meta-review included studies from two review journals only. Although we believe these journals represent the utmost quality and superior methodology, it is still a limited sampling frame. It is possible that the meta-analyses included in these journals represent only one type of review and therefore systematically differ from the population of reviews. We posit, however, that because these journals are some of the most selective in the fields, the problem of publication bias might well be *under*-estimated: The meta-analyses included in these selective journals may include more grey literature than those published in less selective or competitive journals. Future research should attempt to replicate these findings using meta-analyses published across multiple journals in education and psychology.

It is also important to consider that in the current study, the effect size estimates of the difference between published and unpublished studies represent averages of averages. Indeed, the totality of primary studies represented in the meta-review is large ($k = 6{,}392$). It is entirely possible, in fact plausible, that a number of these effect sizes will have inherent bias due to any number of factors. Averaging these effect sizes may balance this bias and render it nominal, but it is nonetheless important to note that we are providing seemingly precise estimates for rather crude approximations. Thus, although we are confident in the overall conclusion that published studies yield larger average effects than unpublished studies, it is difficult to state with great certainty the exact magnitude of the average difference between published and unpublished effect sizes. As with any statistical estimate, a level of uncertainty is inherent in these estimates.

Finally, and most importantly, it is highly probable that the meta-analytic literature, similar to primary literature (Pigott et al., 2013), suffers from outcome reporting biases. In other words, it is likely that meta-analysts test for differences between published and unpublished primary studies but report the results of this test only when the test is statistically significant. We believe this is highly probable because review authors often make such statements. Wood, Lundgren, Ouellette, Busceme, and Blackstone's (1994) moderator results, for example, indicated that unpublished studies actually yielded stronger effect sizes relative to published studies for one outcome. However, the authors conducted several syntheses and tested for differences between published and unpublished studies, yet only reported the moderator results when significant differences were found. As a result, this particular meta-analysis provided a somewhat biased estimate of the difference between published and unpublished studies, albeit in an opposite direction than we hypothesized. Nevertheless, this is one clear example of outcome reporting bias. Meta-analysts, in the future, would do well to follow the reporting practices set forth by primary researchers with regard to outcome reporting.

Future Research

The social sciences need greater awareness of publication bias and its potential negative effects, particularly given the increased role of meta-analyses in informing policy and practice (Ferguson & Brannick, 2012). Yet, as Rothstein et al. (2005) argued, little empirical research has focused on publication bias in the social sciences, relative to the medical sciences, where it has been researched extensively. For instance, we advocate for greater use of cohort studies in the social and behavioral sciences that track the research process of federally-funded primary studies (e.g., Chan et al., 2004). These types of studies can help elucidate where dissemination bias occurs and how to potentially stop it. As for meta-analysis, future research is needed on how to handle the presence of publication bias. One promising option is the use of selection models that account for outcome-reporting bias and adjust parameters accordingly (Hedges & Vevea, 2005). These models are still underdeveloped, however, and offer response to only one type of dissemination bias. Clearly, there is a need for new innovations in meta-analysis methodology that can be used to assess and address the problems associated with publication bias.

Implications

Given the rise in meta-analyses and their broad reach, it should be paramount to ensure that the results of meta-analyses are valid. Policymakers and practitioners rely on meta-analysis results to guide decision-making and decide which programs receive support. Stakeholders rely on results from meta-analyses due to the implicit understanding that they represent the totality of evidence for a given research question. Although it is tempting, and often less time-consuming, to only include studies from peer-reviewed journals in a meta-analysis, researchers must ensure that all available databases, resources, and contacts have been exhausted in the literature search for a meta-analysis.

Publication bias is not only a problem for meta-analysis. The over-represented publication of statistically significant findings has the potential to disrupt the development of new theories and theory testing, research replication, future study planning, and the allocation of resources to research. All of these issues concern the primary researcher, as well as the meta-analyst. Primary researchers should take pains to publish the confirmation as well as the contradiction of theory and program testing.

Conclusions

Evidence is available from methodological reviews to support the existence of publication bias, particularly in the health and medical fields (Rothstein et al., 2005). The current study is one in a growing body of literature that suggests publication bias is also prevalent in the social and behavioral sciences (Pigott et al., 2013; Tanner-Smith & Polanin, 2014). Given the increased awareness of publication bias among primary study authors, meta-analysts, peer reviewers, and journal editors, we may find less evidence of publication bias going forward. Any such improvements henceforth, however, will not counteract the existence of publication bias in decades past. Systematic reviewers and meta-analysts must continue to search for, and include, grey literature in their reviews because of this phenomenon. Peer reviewers and editors should ask for, at a minimum, a clear description of the study selection and sources of literature. At most, this important stakeholder group could demand the inclusion of grey literature. Meta-analytic consumers must be aware of the literature body synthesized. The consumer should view reviews of published literature with a level of skepticism until a review of the entire body of literature is completed.

The results of this meta-review indicated that a failure to include unpublished studies in an education or psychology meta-analysis likely yields a biased overall average effect size estimate. Moreover, simply including one grey literature source, such as dissertations, is not a suitable solution to abate potential publication bias. We hope this research inspires future meta-analysts to consider the myriad forms of publications available and motivates methodologists to investigate this paradigm in the social sciences. For meta-analyses to remain relevant and valuable, review authors simply must include all available literature, regardless of publication status.

References

Meta-analyses included in the review are preceded by an asterisk.

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, *82*, 436-476. doi:10.3102/0034654312458162

*Astill, R. G., Van der Heijden, K. B., Van IJzendoorn, M. H., & Van Someren, E. J. (2012). Sleep, cognition, and behavioral problems in school-age children: A century of research meta-analyzed. *Psychological Bulletin*, *138*, 1109-1138. doi:10.1037/a0028204

Bax, L., & Moons, K. G. (2011). Beyond publication bias. *Journal of Clinical Epidemiology*, *64*, 459-462. doi:10.1016/j.jclinepi.2010.09.003

*Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213-238. doi:10.3102/00346543061002213

*Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, *63*, 69-93. doi:10.3102/00346543063001069

*Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*, 29-58. doi:10.3102/00346543074001029

Becker, B. J. (2005). Failsafe *N* or file-drawer number. In Rothstein, H. R., Borenstein, M., & Sutton, A. J. (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 111-126). West Sussex, England: Wiley & Sons, Ltd.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley.

*Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators-A meta-analysis. *Psychological Bulletin*, *132*, 497-523. doi:10.1037/0033-2909.132.4.497

*Bowman, N. A. (2011). Promoting participation in a diverse democracy a meta-analysis of college diversity experiences and civic engagement. *Review of Educational Research*, *81*, 29-68. doi:10.3102/0034654310383047

*Cantor, J. M., Blanchard, R., Robichaud, L. K., & Christensen, B. K. (2005). Quantitative reanalysis of aggregate data on IQ in sexual offenders. *Psychological Bulletin*, *131*, 555-568. doi:10.1037/0033-2909.131.4.555

Chan, A. W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*, *291*, 2457-2465. doi:10.1001/jama.291.20.2457

Cheung, M. W. L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 429-454. doi:10.1080/10705511.2013.797827

*Cooper, H., Allen, A. B., Patall, E. A., & Dent, A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research*, *80*, 34-70. doi:10.3102/0034654309359185

*Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*, 1-62. doi:10.3102/00346543076001001

*Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitudes. *Review of Educational Research*, *73*, 1-52. doi:10.3102/00346543073001001

*Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, *77*, 113-143. doi:10.3102/003465430298563

Cumming, G. The new statistics. Why and how. *Psychological Science*, *25*, 7-29. doi:10.1177/0956797613504966

*Currier, J. M., Neimeyer, R. A., & Berman, J. S. (2008). The effectiveness of psychotherapeutic interventions for bereaved persons: a comprehensive quantitative review. *Psychological Bulletin*, *134*, 648-661. doi:10.1037/0033-2909.134.5.648

*Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*, 627-668. doi:10.1037/0033-2909.125.6.627

Decullier, E., Lhéritier, V., & Chapuis, F. (2005). Fate of biomedical research protocols and publication bias in France: Retrospective cohort study. *BMJ*, *331*, 19-25. doi:10.1136/bmj.38488.385995.8F

*Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin*, *112*, 106-124. doi:10.1037//0033-2909.112.1.106

*Douglas, K. S., Guy, L. S., & Hart, S. D. (2009). Psychosis as a risk factor for violence to others: A meta-analysis. *Psychological Bulletin*, *135*, 679-706. doi:10.1037/a0016311

*Dush, D. M., Hirt, M. L., & Schroeder, H. E. (1989). Self-statement modification in the treatment of child behavior disorders: A meta-analysis. *Psychological Bulletin, 106*, 97-106. doi:10.1037/0033-2909.106.1.97

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-463. doi:10.1111/j.0006-341X.2000.00455.x

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., ... & Williamson, P. R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS One, 3*, e3081. doi:10.1371/journal.pone.0003081

*Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109-128. doi:10.1037/0033-2909.110.1.109

*Eagly, A. H., Chen, S., Chaiken, S., & Shaw-Barnes, K. (1999). The impact of attitudes on memory: An affair to remember. *Psychological Bulletin*, *125*, 64-89. doi:10.1037/00332909.125.1.64

*Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, *129*, 569-591. doi:10.1037/0033-2909.129.4.569

*Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*, 309-330. doi:10.1037/0033-2909.100.3.309

Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*, 867-872. doi:10.1016/0140-6736(91)90201-Y

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629-634. doi:10.1136/bmj.315.7109.629

*Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203-235. doi:10.1037//0033-2909.128.2.203

*Erel, O., & Burman, B. (1995). Interrelatedness of marital relations and parent-child relations: A meta-analytic review. *Psychological Bulletin*, *118*, 108-132. doi:10.1037/0033-2909.118.1.108

*Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin*, *104*, 226-235. doi:10.1037/0033-2909.104.2.226

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*, 120-128. doi:10.1037/a0024445

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555-561. doi:10.1177/1745691612459059

Fisher, Z. & Tipton, E. (2014). robumeta (Version 1.1) [Software]. Available from: http://cran.r-project.org/web/packages/robumeta/robumeta.pdf

*Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological Bulletin*, *132*, 823-865. doi:10.1037/0033-2909.132.6.823

*Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*, 778-882. doi:10.1037/0033-2909.132.5.778

*Gliessman, D. H., Pugh, R. C., Dowden, D. E., & Hutchins, T. F. (1988). Variables influencing the acquisition of a generic teaching skill. *Review of Educational Research*, *58*, 25-46. doi:10.3102/00346543058001025

*Grabe, S., & Hyde, J. S. (2006). Ethnicity and body dissatisfaction among women in the United States: A meta-analysis. *Psychological Bulletin*, *132*, 622-640. doi:10.1037/0033-2909.132.4.622

*Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. *Psychological Bulletin*, *134*, 460-476. doi:10.1037/0033-2909.134.3.460

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1-20. doi:10.1037/h0076157

Harbord, R. M., Egger, M., & Sterne, J. A. C. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine, 25*, 3443-3457. doi:10.1002/sim.2380

*Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99-136. doi:10.3102/00346543066002099

Hedges, L. V., & Vevea, J. (2005). Selection method approaches. In H. R. Rothstein, M. Borenstein, & A. J. Sutton (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 145-174). West Sussex, England: Wiley & Sons, Ltd.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39-65. doi:10.1002/jrsm.5

Higgins, J. P. T., & Green, S. (2011). Cochrane handbook for systematic reviews of interventions version 5.1. [updated March 2011]. *The Cochrane Collaboration*. Retrieved from http://handbook.cochrane.org/

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539-1558. doi:10.1002/sim.1186

Hopewell, S., Clarke, M., & Mallett, S. (2005). Grey literature and systematic reviews. In H. R. Rothstein, M. Borenstein, & A. J. Sutton (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 48-72). West Sussex, England: Wiley & Sons.

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*, *1*, 1-25. doi:10.1002/14651858.MR000006.pub3

*Hostetter, A.B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*, 297-315. doi:10.1037/a0022128

Ioannidis, J. P. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA*, *279*, 281-286. doi:10.1001/jama.279.4.281

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*, e124. doi:10.1371/journal.pmed.0020124

Ioannidis, J. P., Munafo, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in Cognitive Sciences*, 18, 235-241. doi:10.1016/j.tics.2014.02.010

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*, 245-253. doi:10.1177/1740774507079441

*Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, *126*, 703-726. doi:10.1037/0033-2909.126.5.703

*Johnson, R. E., Chang, C. H., & Lord, R. G. (2006). Moving from cognition to behavior: What the research says. *Psychological Bulletin*, *132*, 381-415. doi:10.1037/0033-2909.132.3.381

*Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*, 376-407. doi:10.1037/0033-2909.127.3.376

*Juffer, F., & Van IJzendoorn, M. H. (2007). Adoptees do not lack self-esteem: A meta-analysis of studies on self-esteem of transracial, international, and domestic adoptees. *Psychological Bulletin*, *133*, 1067-1083. doi:10.1037/0033-2909.133.6.1067

*Kim, J. S., & Quinn, D. M. (2013). The Effects of Summer Reading on Low-Income Children's Literacy Achievement From Kindergarten to Grade 8 A Meta-Analysis of Classroom and Home Interventions. *Review of Educational Research*, *83*, 386-431. doi:10.3102/0034654313483906

*Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*, 616-642. doi:10.1037/a0023557

*Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79-97. doi:10.3102/00346543058001079

Kulik, J. A., & Kulik, C. L. C. (1989). Meta-analysis in education. *International Journal of Educational Research*, *13*, 321-340. doi:10.1037/0003-066X.48.12.1181

*LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin*, *129*, 305-334. doi:10.1037/0033-2909.129.2.305

*Lanaj, K., Chang, C. H., & Johnson, R. E. (2012). Regulatory focus and work-related outcomes: a review and meta-analysis. *Psychological Bulletin*, *138*, 998-1034. doi:10.1037/a0027723

*Lauer, P.A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 275-313. doi:10.3102/00346543076002275

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*, 2-17. doi:10.1002/asi.22784

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, *76*, 286-302. doi:10.1080/03637750903074685

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181-1209. doi:10.1037/0003-066X.48.12.1181

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage Publication.

Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 147-158). New York, NY: Sage Publication.

*Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*, 449-521. doi:10.3102/00346543071003449

*Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423-458. doi:10.3102/00346543066004423

*Lytton, H., & Romney, D.M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, *109*(2), 267-296. doi:10.1037/0033-2909.109.2.267

*Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*, 300-335. doi:10.3102/0034654310377087

*McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424-453. doi:10.1037/0033-2909.126.3.424

*Metzger, S. A., & Wu, M. J. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, *78*, 921-940. doi:10.3102/0034654308323035

*Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, *128*, 638-662. doi:10.1037/0033-2909.128.4.638

*Murray, J., Farrington, D. P., & Sekol, I. (2012). Children's antisocial behavior, mental health, drug use, and educational performance after parental incarceration: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 175-210. doi:10.1037/a0026407

*Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*, 413-448. doi:10.3102/00346543076003413

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159. doi:10.2307/1164923

*Pascoe, E. A., & Smart Richman, L. (2009). Perceived discrimination and health: A meta-analytic review. *Psychological Bulletin*, *135*, 531-554. doi:10.1037/a0016059

*Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, *134*, 270-300. doi:10.1037/0033-2909.134.2.270

*Patall, E. A., Cooper, H., & Robinson, J. C. (2008). Parent involvement in homework: A research synthesis. *Review of Educational Research*, *78*, 1039-1101. doi:10.3102/0034654308325185

*Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research*, *74*, 215-253. doi:10.3102/00346543074002215

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*, 676-680. doi:10.1001/jama.295.6.676

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, *42*, 424-432. doi: 10.3102/0013189X13507104

Polanin, J. R., & Pigott, T. D. (2014). The use of meta-analytic statistical significance testing. *Research Synthesis Methods.* Advanced online publication. doi: 10.1002/jrsm.1124

*Qin, Z., Johnson, D. W., & Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. *Review of Educational Research*, *65*, 129-143. doi:10.3102/00346543065002129

*Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*, 146-174. doi:10.1037/a0025750

*Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, *137*, 131-148. doi:10.1037/a0021705

*Rind, B., Tromovitch, P., & Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, *124*, 22-53. doi:10.1037/0033-2909.124.1.22

*Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, *79*, 3-38. doi:10.3102/0034654308325690

Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine*, *53*(3), 247-271.

*Ross, J. A. (1988). Controlling variables: A meta-analysis of training studies. *Review of Educational Research*, *58*, 405-437. doi:10.3102/00346543058004405

Rothstein, H. R., Borenstein, M., & Sutton, A. J. (2005). *Publication bias in meta-analysis. Prevention, assessment and adjustments.* West Sussex, England: Wiley & Sons.

Rücker, G., Schwarzer, G., & Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine*, *27*, 746-763. doi:10.1002/sim.2971

Schild, A. H. E., & Voracek, M. (2014). Finding your way out of the forest without a trail of bread crumbs: Development and evaluation of two novel displays of forest plots. *Research Synthesis Methods*. Advanced online publication. doi:10.1002/jrsm.1125

*Sedlmeier, P., Eberth, J., Schwarz, M., Zimmermann, D., Haarig, F., Jaeger, S., & Kunze, S. (2012). The psychological effects of meditation: A meta-analysis. *Psychological Bulletin*, *138*, 1139-1171. doi: 10.1037/a0028168

*Seto, M. C., & Lalumiere, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin*, *136*, 526-575. doi:10.1037/a0019700

Simonsohn, U., Nelson, Leif, D., & Simmons, J. P. (2014). *P*-curve: A key to the file drawer. *Journal of Experimental Psychology*, *143*, 534-547. doi:10.1037/a0033242

Smith, M. L. (1980). Publication bias and meta-analysis. *Evaluation in Education, 4*, 22-24. doi:10.1016/0191-765X(80)90004-X

*Smith, T. B., McCullough, M. E., & Poll, J. (2003). Religiousness and depression: evidence for a main effect and the moderating influence of stressful life events. *Psychological Bulletin*, *129*, 614-636. doi:10.1037/0033-2909.129.4.614

*Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics a meta-analysis. *Review of Educational Research*, *81*, 97-128. doi:10.3102/0034654310378174

*Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, *139*, 213-240. doi:10.1037/a0028931

*Steblay, N. M. (1987). Helping behavior in rural and urban environments: A meta-analysis. *Psychological Bulletin*, *102*, 346-356. doi:10.1037/0033-2909.102.3.346

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, M. Borenstein, & A. J. Sutton (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 75-98). West Sussex, England: Wiley & Sons.

*Stice, E., Shaw, H., & Marti, C. N. (2006). A meta-analytic review of obesity prevention programs for children and adolescents: The skinny on interventions that work. *Psychological Bulletin*, *132*, 667-691. doi:10.1037/0033-2909.132.5.667

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435-543). New York, NY: Sage Publication.

*Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, *68*, 277-321. doi:10.3102/00346543068003277

Tanner-Smith, E. E., & Polanin, J. R. (2014). A retrospective analysis of dissemination biases in the brief alcohol intervention literature. *Psychology of Addictive Behaviors*. Advanced online publication. doi:10.1037/adb0000014

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *4*, 13-30. doi:10.1002/jrsm.1091

*Thoresen, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin*, *129*, 914-945. doi:10.1037/0033-2909.129.6.914

*Toosi, N. R., Babbitt, L. G., Ambady, N., & Sommers, S. R. (2012). Dyadic interracial interactions: A meta-analysis. *Psychological Bulletin*, *138*, 1-27. doi:10.1037/a0025767

Torgerson, C. J. (2006). Publication bias: The achilles' heel of systematic reviews? *British Journal of Educational Studies*, *54*, 89-102. doi:10.1111/j.1467-8527.2006.00332.x

*Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*, 352-402. doi:10.1037/a0028446

*Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, *135*, 452-477. doi:10.1037/a0015329

* Van Ijzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and

school performance. *Psychological Bulletin*, *131*, 301-316. doi:10.1037/0033-2909.131.2.301

*Vachon, D. D., Lynam, D. R., & Johnson, J. A. (2014). The (non)relation between empathy and aggression: Surprising results from a meta-analysis. *Psychological Bulletin*, *140*, 751-773. doi:10.1037/a0035236

*Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, *66*, 323-340. doi:10.3102/00346543066003323

*Vitaliano, P. P., Zhang, J., & Scanlan, J. M. (2003). Is caregiving hazardous to one's physical health? A meta-analysis. *Psychological Bulletin*, *129*, 946-972. doi:10.1037/0033-2909.129.6.946

*Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*, 249-268. doi:10.1037/0033-2909.132.2.249

*Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin*, *132*, 132-149. doi:10.1037/0033-2909.132.1.132

*Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. (1994). Minority influence: A meta-analytic review of social influence processes. *Psychological Bulletin*, *115*, 323-346.

*Wood, W., & Quinn, J. M. (2003). Forewarned and forearmed? Two meta-analysis syntheses of forewarnings of influence appeals. *Psychological Bulletin*, *129*, 119-138. doi:10.1037/0033-2909.129.1.119

*Wood, W., Wong, F. Y., & Chachere, J. G. (1991). Effects of media violence on viewers' aggression in unconstrained social interaction. *Psychological Bulletin*, *109*, 371-383. doi:10.1037/0033-2909.109.3.371

Table 1

Included Meta-Analyses' Characteristics

First Author	Date	Source	Synthesis Description	Outcome; Effect Size; Model	Funded; Bias	Grey Literature	Total Studies; Unpublished
Astill	2012	PB	Relationships between sleep, cognitive performance, and behavioral problems	P, A; C; R	Y; T, FP	D; Co	86; 17
Bangert-Drowns	1991	RER	Intervention effects of feedback on achievement	A; S; N	N; M	D; G	58; 28
Bangert-Drowns	1993	RER	Intervention effects of word processors on achievement	A; S; N	N; M	D	20; 14
Bangert-Drowns	2004	RER	Intervention effects of write-to-learn programs on achievement	A; S; F	Y; M	G	48; 13
Bösch	2006	PB	Difference between human and number generators	P; S; B	Y; T	G	355; 78
Bowman	2011	RER	Correlation between college diversity and civic engagement	P; C; R	N; T	D	28; 5
Byrnes	1999	PB	Gender differences in risk taking	P; S; N	N; M	D	150; 15
Cantor	2005	PB	Difference between sexual offenders and non-offenders on IQ	P; S; R	Y; M	D; G	161; 22
Cooper	2003	RER	Intervention effects of calendar modifications	A; S; B	Y; M	D; Co; G	41; 33
Cooper	2006	RER	Relationship between time on homework and academic achievement	P, A; C; F	Y; T	D; Co; G	35; 18
Cooper	2010	RER	Intervention effects of full day vs half day kindergarten	A; S; B	N; T	D; G	40; 33
Cornelius-White	2007	RER	Relationship between teacher-student dyad and achievement	A; C; R	Y; M	D; BC; B; C; G	118; 40
Currier	2008	PB	Intervention effects of psychotherapy on bereavement	P; S; R	N; T	D	64; 16
Deci	1999	PB	Experimental effects of rewards on intrinsic motivation	P; S; N	Y; M	D	70; 21
Dindia	1992	PB	Sex differences in self disclosure	P; S; N	N; M	D	200; 51
Douglas	2009	PB	Relationship between psychosis and violence	P; C; R	N; F	D	166; 6
Dush	1989	PB	Intervention effects of self-statement modification on child behavior disorders	P; S; N	N; M	D; Co	48; 16

Eagly	1986	PB	Sex differences in aggressive behavior	P; S; N	Y; M	D	50; 8
Eagly	1991	PB	Physical attractiveness differences on life outcomes	P; S; N	Y; M	D; G	76; 23
Eagly	1999	PB	Experimental effects of attitudes on memories	P; S; F	Y; M	D; Co; G	62; 18
Eagly	2003	PB	Women vs men in leadership style	P; S; R	Y; M	D; B; G	44; 28
Elfenbein	2002	PB	Relationship between cultural specificity and emotion recognition	P; C; N	Y; FD	D; Co; B; BC; G	43; 17
Erel	1995	PB	Marital relations and parent-child relationship	P; C; N	Y; M	D; BC	69; 22
Feingold	1988	PB	Relationship between partner attractiveness: samples	P; C; N	N; M	D; Co; B	19; 6
Frattaroli	2006	PB	Experimental effects of disclosure	P; S; B	Y; FD	D; Co; G	146; 70
Glasman	2006	PB	Relationship between attitudes and future behavior	P; C; B	Y; M	D; G	42; 8
Gliessman	1988	RER	Intervention effect of teaching skills on achievement	A; S; F	N; M	G	26; 17
Grabe	2006	PB	Difference between races on body dissatisfaction	P; S; R	N; M	D	93; 16
Grabe	2008	PB	Experimental effects of media on body image	P; S; R	Y; M	D; G	20; 3
Hattie	1996	RER	Whether certain interventions improve academic performance, study skills, or affect	A; S; N	Y; M	D; BC; G	51; 6
Hostetter	2011	PB	Gestures vs no gestures in communication	A; S; R	N; F	D; Co	63; 15
Jaffee	2000	PB	Sex differences in moral orientation	P; S; N	N; M	D; Co; B; G	114; 44
Johnson	2006	PB	Experimental effect: personal goal vs accessibility of goal-related attainment	A; S; F	N; T	D; Co	27.5; 7
Judge	2001	RER	Correlation between job satisfaction and performance	P; C; N	N; M	D	312; 92
Juffer	2007	PB	Difference between adopt vs non-adopt	P; S; R	Y; T	B; G	88; 44
Kim	2013	RER	Effect of summer reading on reading comprehension	A; S; R	N; T, FD	D; BC; G	35; 24
Koenig	2011	PB	Relationship between leadership stereotypes and masculinity	P; C; R	Y; T, FP	D; G	50; 12
Kulik	1988	RER	Intervention effects of immediate feedback	A; S; N	N; M	D	27; 3
LaFrance	2003	PB	Difference of sexes	P; S; F	Y; F	D; B; BC; Co	418; 126
Lanaj	2012	PB	Antecedents and consequences of regulatory	P; C; R	N; M	D; Co	125; 55

			focus				
Lauer	2006	RER	Intervention effects of out of school programs	A; S; R	Y; M	D; Co; G	26; 22
Lou	1996	RER	Intervention effects of grouping on academic achievement	A; S; N	Y; M	D; Co; G	105; 17
Lou	2001	RER	Intervention effects of grouping on computer technology learning	A; S; N	Y; M	D; Co; G	100; 48
Lytton	1991	PB	Parental differential towards boys and girls	P; S; N	Y; M	D; Co; G	40; 8
Marulis	2010	RER	Intervention effects of vocab interventions	A; S; R	Y; F	D; G	62; 15
McClure	2000	PB	Sex differences in facial expression processing	P; S; R	N; FD	D	80; 21
Metzger	2008	RER	Relationship between teacher perceiver interview and teacher quality	A; C; B	Y; FP	D; G	45; 43
Mor	2002	PB	Relationship between attention and negative affect	P; C; N	N; F	D; G	104; 20
Murray	2012	PB	Difference between children with and without incarcerated parents	P, A; C; R	Y; T	D	40; 20
Nesbit	2006	RER	If using concept maps increased knowledge retention	A; S; R	Y; M	D	40; 12
Pascoe	2009	PB	Relationship between discrimination and health	P; C; B	Y; T	D; Co; G	192; 19
Patall	2008	RER	Relationship between parental involvement and achievement	A; C; B	N; T	D; Co; G	22; 11
Patall	2008	PB	Experimental effects of choice on intrinsic motivation	P; S; B	N; T	D	46; 18
Penny	2004	PB	Intervention effects of consultation on teaching effectiveness	A; S; F	N; M	D	11; 4
Qin	1995	RER	Intervention effect	A; S; N	N; M	D; B	59; 10
Rhodes	2011	PB	Experimental effects of delayed judgment on metacognitive accuracy	A; S; F	N; FD	D; G	112; 22
Rhodes	2012	PB	Facial recognition of one's own age vs. other ages	P; S; R	N; FD	Co; G; BC	46; 4
Rind	1998	PB	Adjustment of victims of child sexual abuse	P; C; N	N; M	D	62; 23
Ritter	2009	RER	Intervention effects of volunteer tutor programs	A; S; R	Y; T	D; G	14; 5
Ross	1988	RER	Intervention effects of programs on students	A; S; N	N; M	D; G	61; 20
Sedlmeier	2012	PB	Intervention effect of meditation on emotionality and attention	P; S; F	N; FD, FP	D; BC	164; 39
Seto	2010	PB	Difference between male sex offenders and others	P; S; R	N; M	D; BC; Co; G	28; 17

Smith	2003	PB	Relationship between religiousness and depression	P; C; R	Y; FD, FP	D; Co	147; 36
Sosa	2011	RER	Intervention effects of computer-assisted learning	A; S; B	N; FD	D; Co	45; 13
Sowislo	2013	PB	Relationship between self-esteem and depression	P; C; R	Y; FP	D	77; 7
Steblay	1987	PB	Relationship between helping behavior and living environment	P; C; N	Y; FD	D; Co	35; 4
Stice	2006	PB	Intervention effects of prevention programs on obesity	P; S; R	Y; M	D	64; 2
Swanson	1998	RER	Intervention effects of programs for learning disabled students	A; S; F	Y; F	D	180; 25
Thoresen	2008	RER	Correlation between job attitude and performances	P; C; R	N; F	D	37; 7
Toosi	2012	PB	Interracial interactions and attitudes, emotional states	P, A; C; R	Y; T, FP	D; G	81; 9
Uttal	2013	PB	Intervention effects on spatial skills	A; S; R	Y; FD, FP	D; BC; B; Co; G	206; 111
Vachon	2013	PB	Relationship between empathy and aggression	P; C; B	N; FP	D	103; 31
Van den Bussche	2009	PB	Before vs after reaction times for different stimulus	P; S; R	N; T	D; G	28; 4
Van Ijzendoorn	2005	PB	Difference between adopted and non-adopted kids	A; S; R	Y; T	D; B; BC; G	36; 4
Veenman	1996	RER	Multi-grade vs single-grade class effects	A; S; R	N; M	D; B; G	40; 29
Vitaliano	2003	PB	Difference between caregivers and non-caregivers on physical health	P; S; R	Y; T	D	17; 2
Webb	2006	PB	Intervention effect on intention-change	A; S; R	N; M	D; G	47; 5
Weisz	2006	PB	Intervention effects of psychotherapy on depression	P; S; R	Y; M	D	35; 8
Wood	1991	PB	Experimental effects of media violence on aggression	P; S; N	Y; M	D; Co	14; 4
Wood	1994	PB	Experimental effects of minority influence on social processes	P; S; N	Y; M	D	23; 6
Wood	2003	PB	Experimental effects of forewarning on self-image	P; S; R	Y; M	D	20; 2

Note. First author represents the meta-analyses' first author's last name; RER = *Review of Educational Research*; PB = *Psychological Bulletin*; A = Achievement outcome; P = Psychological outcome; S = Standardized mean-difference; C = Correlation coefficient; N = Not reported; F = Fixed-effect; R = Random-effects; B = Both fixed- and random-effects; Y = Funded; M = Moderator only; T = Trim and fill; FD = File-drawer; FP = Funnel plot; D = Dissertation; G = General unpublished; B = Book; BC = Book chapter; Co = Conference proceeding.

Table 2

Grey Literature Included in the Sample of Meta-Analyses

Publication Types of Studies included in the Meta-Analysis	Meta-Analyses with Pub. Type (%)	Number of Studies Included in Meta-Analysis (Mean, SD)	Minimum, Maximum	Average % of Total
Journal Articles	81 (100)	55.74 (56.74)	2, 292	68.5
Unpublished Literature	81 (100)	22.19 (23.11)	2, 126	31.5
Dissertations/Theses	64 (79.0)	13.43 (18.42)	0, 92	18.0
Conference Abstracts	26 (32.1)	2.34 (8.20)	0, 68	2.80
Books (General)	8 (9.87)	0.56 (2.96)	0, 25	0.52
Book Chapters	12 (14.8)	0.82 (3.40)	0, 9	0.81
General Unpublished	45 (55.6)	5.03 (2.96)	0, 45	9.40

Note. k = 81; Unpublished literature row represents the totality of unpublished studies; General unpublished is used when review authors did not provide grey literature types or to represent all other types of unpublished literature (i.e., unpublished datasets, reports, etc.).

Table 3

Meta-Regression Analysis

Model	I	II	III	IV	V	VI	VII
Characteristics							
Effect size type	-0.01						-0.14
	[-0.83, 0.81]						[-0.96, 0.68]
Dissertations only		-0.77*					-0.83*
		[-1.51, -0.03]					[-1.63, -0.03]
Percent unpublished			-0.03				-0.19
			[-1.57, 1.51]				[-.1.83, 1.46]
Funded				0.27			0.18
				[-0.42, 0.95]			[-0.49, 0.85]
Date of publication					-0.03		-0.04
					[-0.08, 0.01]		[-0.09, 0.01]
Q-between conversion						0.47	0.60
						[-0.42, 1.36]	[-0.34, 1.53]
Intercept	0.80*	1.03*	0.81*	0.65*	0.75*	0.71*	0.94*
	[0.06, 1.54]	[0.60, 1.45]	[0.18, 1.44]	[0.26, 1.05]	[0.42, 1.08]	[0.33, 1.10]	[0.13, 2.01]
R^2	5.38	15.16	1.97	11.01	0.01	0.01	5.16

Note. Number of effect sizes = 147; Number of meta-analyses = 81; Results estimated from mixed-effect meta-regression models with robust standard errors; Numbers in brackets represent 95% confidence intervals; Effect size type (0 = Correlation coefficient, 1 = Standardized mean-difference); Dissertations only (1 = Dissertations only); Funded (1 = Received funding); Q-Between conversion (1 = Converted from *Q*-between tests); * $p < .05$.

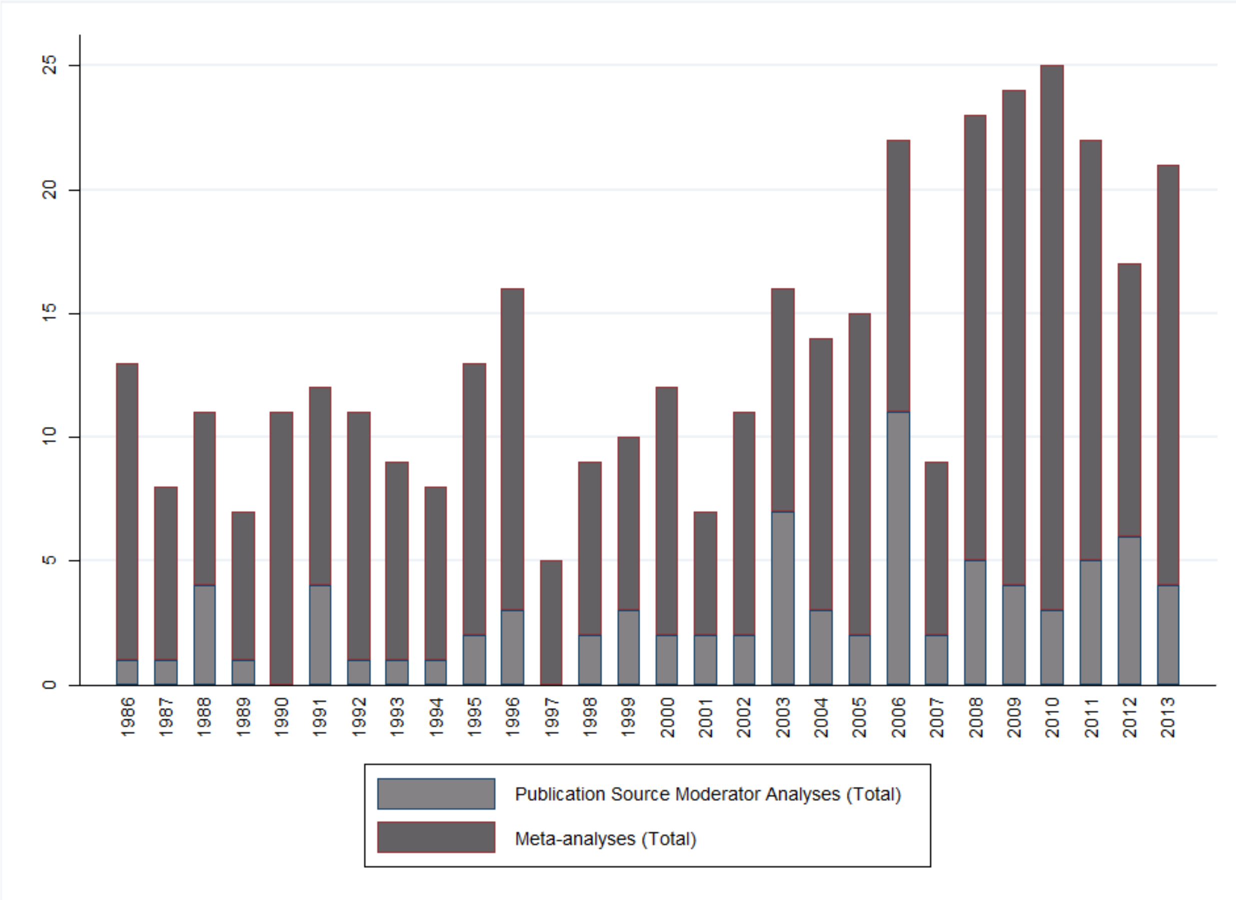
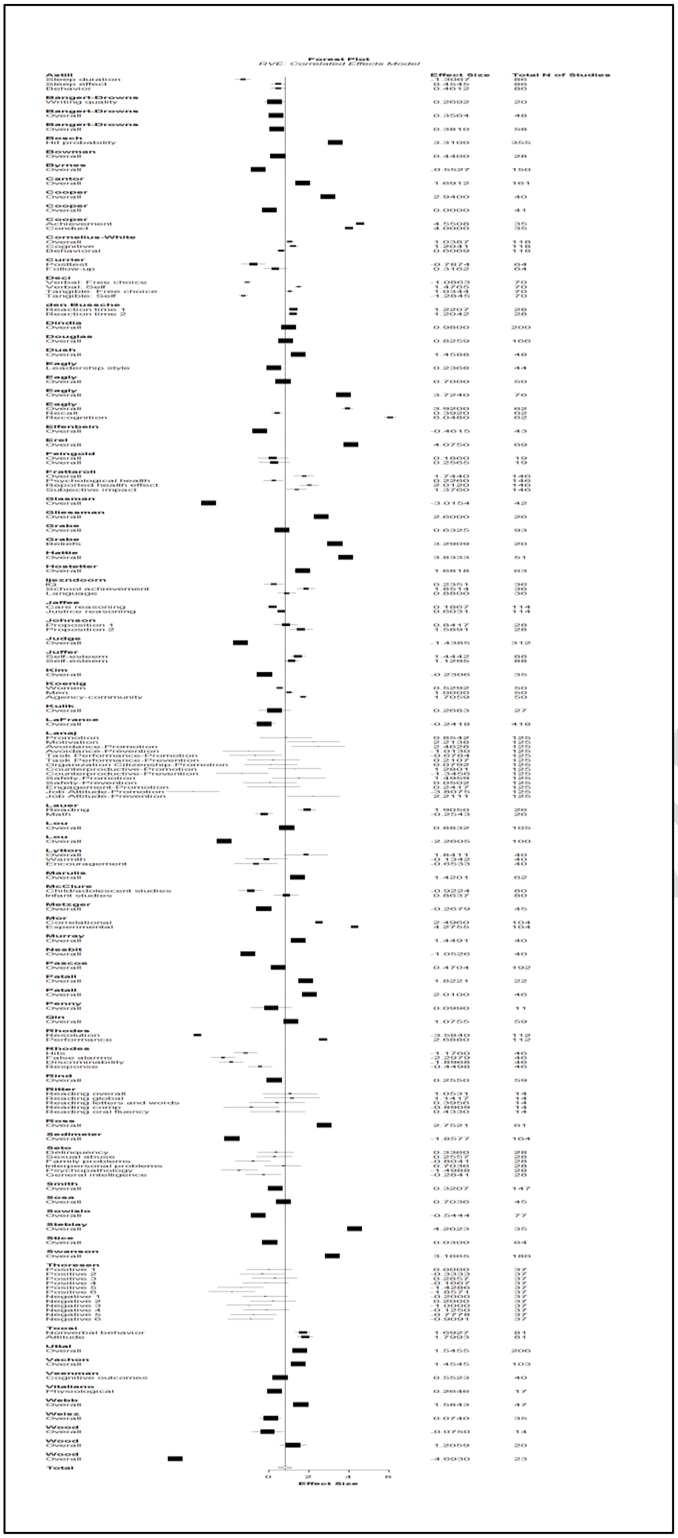


Figure 1. *Number of Moderator Analyses Conducted Per Year.*



Online Supplemental Figure S1: Forest Plot of all Effect Sizes Calculated.